COMPUTATION METHOD AND SYSTEM FOR UNSTRUCTURED TEXT DATA

ABSTRACT

The present invention provides a computation method and system for unstructured text data, and relates to the technical field of natural language processing. The method includes: performing multi-level processing on the input unstructured text data, dynamically adjusting word segmentation granularity according to word frequency distribution, constructing a part-of-speech co-occurrence matrix in combination with context semantic information, extracting entity information, and fusing the part-of-speech co-occurrence matrix with the entity information to generate a hierarchical semantic tag sequence; extracting feature representation by feature extraction units of different convolution kernel size, and establishing an associated weight matrix by computing cosine similarity between different semantic hierarchies; constructing a semantic enhancement vector based on entity information, performing adversarial training to obtain a multi-modal semantic feature matrix; and computing semantic similarity of fusion feature vectors, setting difficulty weights according to structural complexity, consistency and ambiguity of class clusters, sorting and then inputting into a classifier, and performing iterative optimization on the classification result, so as to obtain a computation result.

[Fig. 1]